Figure S5. Gene selection frequency across different partitions of the TCGA samples. In machine learning, gene selection instability is an issue, which means different sets of genes can be selected based on different subsets of training samples used. We partitioned the entire TCGA samples downloaded into the training and test subsets 100 times. In each iteration, we retained the top genes up to 40 that were selected by the algorithm. (In case that less than 40 genes were selected, we retained the actual genes.) Gene frequency was then computed across the 100 iterations. The 16 genes are among the top ones that are most frequently selected.